情境本体驱动的多源知识融合框架*

■ 唐旭丽[1,2] 张斌[2,3,4] 傅维刚[1,2]

[1] 武汉大学信息资源研究中心 武汉 430072 [2] 武汉大学大数据研究院 武汉 430072
[3] 武汉大学国家文化发展研究院 武汉 430072 [4] 武汉大学中国传统文化研究中心 武汉 430072

摘要：[目的/意义]情境建模是解决信息泛滥、信息过载、实现信息按需服务的重要手段，目前已有的知识库构建和知识融合方法普遍忽略了情境信息，阻碍了知识库的实际应用，降低了知识服务的效率和效果。[方法/过程]综合考虑环境情境、个人情境和领域本体三个方面，提出一种情境本体驱动的多源知识融合框架，并以此框架融合生成基于情境的药物不良反应知识库 ConADR Ontology。在本框架的指导下，以药物不良反应的知识库构建为例，半自动实现情境本体模式层的构建和数据层的扩充；并以情境本体作为中介本体，实现情境本体，药物不良反应领域本体 ADReCS 和人类疾病领域本体 Disease Ontology 间的融合；最终在此基础上实现基于 SPARQL 的案例查询。[结果/结论]实例验证表明，本框架具有一定的可行性，对知识库的建设和应用具有理论性指导和参考价值。

关键词：情境本体 多源知识融合 知识库构建 药物不良反应
分类号：G203
DOI：10.13266/j.issn.0252-3116.2018.22.013

大数据时代，数字化信息规模呈现指数级增长，其类型复杂、种类繁多、来源广泛，严重阻碍了用户获取有用信息的进程。如何从庞杂的多源异构数据中抽取有价值的信息并提供高效的知识服务，成为亟需解决的问题。知识库为用户提供经过提炼和验证的高质量的知识单元，成为大数据与知识服务之间的桥梁，为最终高效知识服务提供了有力的保障[1]。

目前，已有知识库大都以本体技术为基础，针对某个领域具体的问题，将领域理论知识、事实数据以及经验描述等依据语义关联有机地组织成知识集群[1]。但是，这些本体知识库的构建过程普遍忽略了情境信息[2-4]。情境建模是知识服务的重要组成部分，情境信息有效地揭示了知识与任务需求间的相关性，结合情境信息向用户提供适合当前情形的信息和服务，是解决信息泛滥、信息过载，实现信息按需服务的重要手段[4]。笔者综合考虑情境因素，并充分利用成熟领域本体的完善的概念体系和逻辑架构，从环境情境、个人情境和领域本体三个方面，提出了一种情境本体驱动的多源知识融合框架，并以此框架为指导，以药物不良反应知识库构建为例，研究了情境本体构建以及知识融合，对知识库的建设和应用提供一些理论性的指导和参考价值。而且，笔者构建及融合生成的基于情境的药物不良反应知识库，为医学智能信息咨询和知识服务提供了语义支撑。

1 相关文献述评

1.1 情境本体和知识服务

知识服务是将用户需求和情境融入用户解决问题的过程中，从而有效支持知识应用和创新服务。其中，情境又称上下文，是指任何可以用来刻画一个实体特征的信息，包括时间、地点、温度、用户等[5]。目前，知识服务处在大数据的环境下，主要面临两方面的挑战：①如何更精确的定义用户需求；②如何从各类多维、异构的数据源中提取有效信息、组织成知识，并支持用户决策。情境建模通过对用户环境和用户资料进行描述，可以协助我们深入理解用户需求，并对无关信息进

* 本文系国家自然科学基金重点国际(地区)合作研究项目“大数据环境下的知识组织与服务创新研究”(项目编号：71420107026)和国家自然科学基金国际(地区)合作与交流项目“基于慢病知识管理的智慧养老平台研究”(项目编号：71661167007)研究成果之一。
作者简介：唐旭丽(ORCID：0000-0002-1656-3014)，博士研究生，E-mail：xulitang@whu.edu.cn；张斌(ORCID：0000-0002-5591-7874)，讲师，博士后；傅维刚(ORCID：0000-0003-4682-696X)，博士研究生。
收稿日期：2018-03-04 修回日期：2018-08-04 本文起止页码：109-117 本文责任编辑：徐健

行过滤,提供精准化的知识服务。

目前,基于本体的情境模型,在提供清晰的情境概念定义的同时,描述了概念间的语义关系,被认为是非常有效的情境建模方法,备受推崇[6-7]。研究者提出了很多情境本体建模的普适性框架,X. H. Wang 等[8]针对普适计算情境,提出了一种基于 OWL 的情境本体 CONON(CONtext Ontology),该本体采用分层结构对 4 种情境实体——位置、活动、用户和设备进行概念拓展,强调模型的可扩展性,并且支持基于情境的逻辑推理;N. Aloui 等[9]为了实现在学习内容中加入注释文档实现情境关联,提出了分层结构的广义情境本体模型,对 6 种学习相关的情境实体——用户、活动、环境、设备、学习重用和学习目标,进行自上而下的迭代建模,这些模型在简洁性、可扩展性、可施行性、Jena 兼容性等角度各有优缺点,但是充分考虑了用户所处的环境、条件、背景等因素,很好地满足了知识服务对用户需求的准确性、可靠性、自适应和推理性等方面的要求,这些模型框架也被广泛地借鉴到各种知识服务之中;李枫林等[2]通过构建基于情境的高血压药物本体,实现了高血压药物的个性化推荐;潘旭伟等[4]提出了一个以用户情境本体为核心的自适应个性化信息服务体系框架,并基于该框架实现了情境感知的个性化信息检索;周莉等[3]针对用户在电子商务中获取个性化商品信息困难的问题,提出以多维分层用户情境本体为核心的个性化商品信息服务框架;M. Sohn 等[7]针对智能家电应用,提出了一种基于模糊论和案例推理的情境本体的个性化服务框架,针对传感器采集的动态数据流,将数值数据转换成为语义词,如“高”“中”“低”等。

但是,目前情境本体的构建只考虑情境信息,普遍上忽略了与领域相关的隐性的背景语义。成熟的领域本体作为某一学科领域中概念及其关系的可复用的术语词汇[10],可以实现对情境本体隐含语义的补充。而且,这些本体大都具备完善的概念体系和逻辑架构,可以通过知识融合的方式,扩展本体的语义互联和可操作性,从而实现将某领域零散的知识聚焦到面向特定领域实际问题的知识库。

1.2 知识融合

知识服务的核心要素是为用户提供经过精炼和验证的高质量的知识单元,这些知识单元一般有多种来源,需要进行处理[11]。知识融合通过对多来源、跨领域、大规模的异构数据进行实时、准确、智能的分析和提取,通过先进的知识处理技术实现知识单元的揭示、共享、关联和发现,进而得到融合后的新知识或新的解决方案,从而达到知识服务的目标。根据 A. Smirnov 的总结,知识融合的一大目标就是产生新的知识,将多个知识源集成为一个新的数据对象,即构建知识库[12]。周宇和欧石燕[13]认为将某领域零散的知识聚焦到面向特定领域实际问题的知识库,可以更好地支撑用户决策。

A. Preece 等[14]把知识融合的过程划分为三个步骤:知识定位、知识转换和知识融合。其中,找到高质量的知识源对知识服务的质量至关重要。刘晓娟等[11]发现知识融合系统的共性之一是使用成熟的领域本体。领域本体是特定领域概念及其关系的可复用的术语词汇[10],而且大都具备完善的体系。但是,领域本体因为是不同的组织和机构根据自身的需求和理念构建而成,充满了异构性和多样性[15]。目前存在的本体结构的文件类型主要有 OWL、RDF、Turtle、OBO、XML、数据库等,通过知识融合可以消除概念的歧义,剔除冗余和错误概念,从而确保知识的质量[16]。研究者对知识融合提出了很多的方法。王丽伟等[10]提出了基于映射乔接点的多领域本体映射和分类聚合理论模型,并实现了 RxNorm 与 NDF-RT 两种药物本体之间的映射;王海栋等[17]提出了一种基于置信度理论的网络知识融合系统模型,弥补了传统知识融合系统在不确定性处理上的不足,并使用反馈自适应机制自动校正置信度因子以避免初始置信度设置的主观性。以上领域本体融合的形式各异,但从方法上而言,主要是通过实体链接和知识合并的方式实现知识融合,通过反馈机制对知识质量进行控制。这种思路对本文具有很好的借鉴作用。

2 情境本体驱动的多源知识融合框架

情境本体在形式上可为多个来源的领域知识的融合提供映射桥接点,显式声明出各个来源知识的内在联系;在内容上可有效地规范表示情境信息,揭示出实际任务的个性化需求。情境本体通过形式和内容的结合,为融合多源领域知识、匹配个性化需求、提供个性化知识服务奠定了基础。笔者参照 A. Preece 等[14]提出的 KRAFT 知识融合系统架构和王海栋等[17]提出的知识融合系统,从知识资源、知识转换和知识融合三个角度,提出一种情境本体驱动的多源知识融合框架(见图 1)。

知识资源层是知识融合的数据基础,包含一切与知识融合过程相关且未经处理的客观信息来源总和。

这些资源包括自传感器的实时环境数据、在社交网站采集的用户个人资料以及相关成熟的领域知识库文件等。从结构上而言，这些数据一般为非结构化的文本、半结构化的网页或者结构异构的本体等，需要经过进一步的数据处理，才能被接下来的知识融合过程所用。

知识转换层通过知识转换器对不同来源的知识进行抽取、分析、转换，将未经处理的知识资源转换成统一的知识表示，从而消除知识资源的异构性。情境本体驱动的知识融合过程需要利用原始知识资源构建情境本体，这一过程是典型的数据驱动的建模过程，根据数据及情境需求，人工构建情境概念模式层，并通过抽取网站及传感器数据，并使用一些模糊函数的方法进行数据处理后，采用算法自动对情境数据层进行扩充，通过必要的数据质量检查来控制情境本体的质量，当新的数据源加入时，可以重复此过程进行情境本体模式层和数据层的更新。另外，对于异构的领域本体，如 OWL、RDF、Turtle、OBO、XML 等格式，需要统一转换为 OWL 模型。

知识融合层是融合系统的核心，主要包括融合算法、融合规则库、融合知识库三部分。情境本体驱动的知识融合过程参考 KRAFT，在解决语义异质的问题时引入了共享本体概念，为问题领域的词汇的语义说明提供一个共享词汇表，从而实现多个来源的本体之间的映射，本文中的情境本体恰好扮演了共享本体的角色。情境本体为所有领域本体提供映射乔接点，根据融合规则库和知识融合算法，可以实现各相关领域本体的集成和综合，融合形成新的知识，并将最终结果存储到融合知识库中。在融合知识库的基础上，可以实现本体数据间的共享和互操作问题，也可以通过一些算法实现知识发现和数据质量检查。

3 药物不良反应的知识融合研究

作为主要的公共健康问题，药物不良反应（Adverse Drug Reaction，ADR）已成为发病率和死亡率的首要原因之一[18]，严重威胁着患者的健康、遵医行为和医疗花费[19]。尽管 ADR 在临床试验期间被监测，但对样本大小和研究人群的实际限制意味着在批准使用之前，无法检测到药物的所有不良反应事件。目前，针对药物的持续预警，形成了 ADR 自发报告系统，如美国的 FDA 不良事件报告系统；以及在此基础上构建了一系列本体知识库，如 SIDER、ADReCS、IntSide 等。但是，药物不良反应具有明显的不确定性、情境依赖性和个体差异性。针对个体患者，药物不良反应的发生与

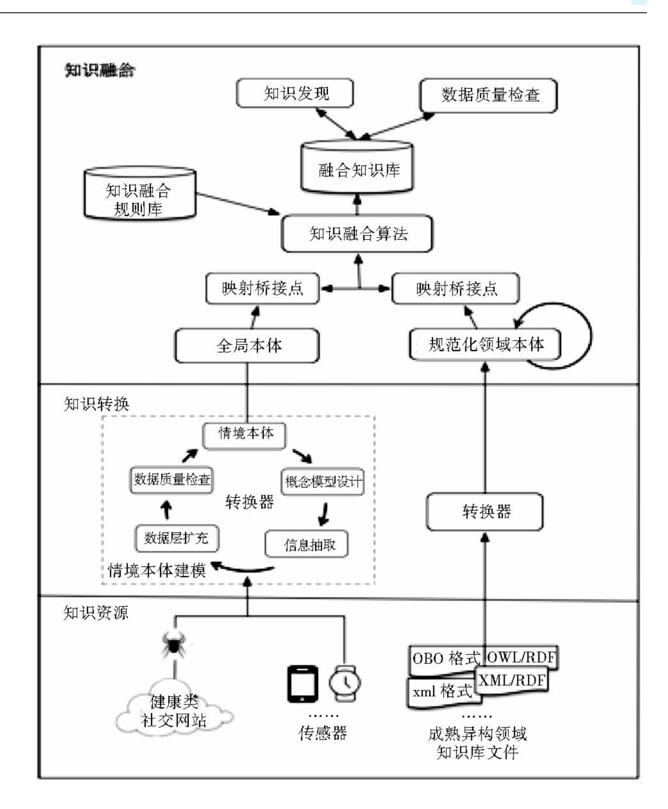


图 1　情境本体驱动的知识融合框架

性别、年龄、门诊、住院、给药方式、历史患病以及合并用药等情境因素息息相关[20]。因此，如何基于情境，融合已有案例和多源本体知识库，为用户提供个性化的药物不良反应案例查询，使其规避药物不良反应的风险，成为亟待解决的问题。笔者以药物不良反应知识库构建为例，演示情境本体驱动的多源知识融合模型构建方法以及查询应用。

3.1　知识资源

3.1.1　情境数据资源　在线患者病例数据共享平台（PatientsLikeMe）是一个典型的病友社交平台，通过互助式在线问诊，患者在平台上主动记录、交流和分享自己的健康问题和治疗情况，寻找相似病例进行参考。PatientsLikeMe 平台发布于 2004 年，由麻省理工学院的工程师 J. Heywood 与其三弟联合创办。目前，社区已拥有 40 多万的会员，涉足病种超过 2 500 个，成为全美最大的健康数据共享平台。平台结构化数据分类较为细致，当查找一个病种时，可以看到所有与该病种相关的病友信息，包括他们的治疗方案、医生开具的用药计划，甚至副作用人群等。

网站汇聚了大量珍贵的病人情境数据，笔者根据研究需要，从 PatientsLikeMe 平台上抓取 235 个用户，包括用户的个人基本资料、历史病例数据、用药情况、药物评价、个人情绪波动等情境信息，共 620 条情境数据。

3.1.2 选取的领域本体

（1）ADReCS（http://bioinf.xmu.edu.cn/ADReCS/index.jsp）是由厦门大学生物信息学辅助药物开发研究组构建的一个关于药物不良反应的知识库，该知识库实现了药物不良反应术语的标准化以及概念层次分类，被国际上广泛使用。该知识库的药物及不良反应数据抽取开源的医疗知识库如 DailyMed、MedDRA、SIDER2、DrugBank、PubChem、UMLS 等，当前包含 21 237个药物不良反应记录，其中有 137 619 个药物和药物不良反应关系对。知识库可在其官网中下载获取，格式为 XML 格式，其中，概念间的层次关系由一组特殊的编码（xx.xx.xx.xxx）来体现。

（2）Disease Ontology（http://www.disease-ontology.org）是由美国马里兰大学医学院提供的一个标准的人类疾病知识库，知识库通过将疾病本体术语交叉映射到医学主题词表 MeSH（Medical Subject Heading Terms）、国际疾病分类法 ICD（International Classification of Diseases）、美国国立癌症研究所叙词表 NCIT（The National Cancer Institute's Thesaurus）、国际系统医学术语集 SNOMED（The Systematized Nomenclature of Medicine）和在线人类孟德尔遗传数据库 OMIM（Online Mendelian Inheritance in Man）等知识库中来达到疾病和医学术语的整合目的，知识库存储为标准的开放生物医学本体论 OBO（Open Biomedical Ontologies）格式，具有很好的可扩展和重用功能。

3.2 知识转换

3.2.1 情境本体的构建

定义 1（情境本体）设情境本体为 $G=<G_s,G_d,R>$。其中，G_s 为模式层，G_d 代表数据层，R 代表两者之间的对应关系。

（1）模式层 $G_s=<N_s,P_s,E_s>$。其中，N_s 表示位置、活动、用户、设备、问题、方案和领域相关实体概念集；P_s 表示关系属性集合，包含：父类关系（subClassOf）、相等关系（equivalentClass）、因果关系（causeClass）以及用户自定义的用来描述概念的相关性质，如药物概念的领域属性；E_s 表示概念节点间的关系集合，$E_s \subseteq N_s \times P_s \times N_s$。

（2）数据层 $G_d=<N_d,P_d,E_d>$ 表示模式层的实例化。其中，N_d 表示实例集合（N_i）或者字面值集合（literals）；P_d 表示属性集合，包括对象属性集合（当 $N_d=N_i$）和数值属性集合（当 $N_d=$ literals）；E_d 表示数据层节点间的关系集合。

（3）G_s 和 G_d 间的关系 $R=\{(\text{instance}, \text{rdf:type}, \text{class}) \mid \text{instance} \in N_i, \text{class} \in N_s\}$。

依据情境本体的定义，在模式层设计时，笔者选取位置、活动、用户、设备、问题和解决方案 6 个上层核心概念，其中，位置、活动、用户、设备参考 CONON 模型[8]；问题与解决方案是综合考虑情境本体对知识服务，如智能咨询服务、用户画像等的支持功能以及在问题发现，如预警上的支持。问题包含用户直接提问或搜索的显性问题以及系统在情境案例本体之上进行挖掘、推断出的用户可能存在的隐性问题两部分。解决方案是通过在情境案例知识库上进行分类、归纳、关联、推理得到的问题解决路径。综合考虑药物不良反应的领域特殊性，笔者对情境本体上层概念进行细化，扩展出药物、药物使用、药物评价、疾病等类，并通过 protégé 构建情境本体的类以及属性层次结构，本体为标准的 OWL/RDF 格式。如图 2 所示，a 是情境本体概念图，b 是情境本体的类，c 是情境本体的对象属性，d 是情境本体的数值属性。

在数据层扩展时，笔者采用 Java 编程的方式，对从 PatientsLikeMe 平台上抓取的用户情境数据进行基本的结构化抽取和规范化表示，并进一步使用 jena 填充情境本体的实例。其中，针对个人信息数据中存在的数值型数据和传感器采集的动态数据流，采纳具有较高计算性能的梯形模糊隶属函数[7,21]，构建用于模糊化的规则，基于 Jena 中自带的规则引擎——GenericRuleReasoner，对数值型数据进行模糊化。

定义 2：模糊函数 $\mu_{i\in A}(X)$：

$$\mu_{i\in A}(X)=\begin{cases} 0, & X\leqslant \alpha_i \\ \dfrac{X-\alpha_i}{\beta_i-\alpha_i}, & \alpha_i<X\leqslant \beta_i \\ 1, & \beta_i<X\leqslant \gamma_i \\ -\dfrac{X-\delta_i}{\delta_i-\gamma_i}, & \gamma_i<X\leqslant \delta_i \\ 0, & \delta_i<X \end{cases}$$

其中，X 为需要进行转换的数值型问题情境，A 是模糊后的语义词属性集合，$\mu_{i\in A}(X)$ 即数值变量 X 在语义词 $i\in A$ 情境下的隶属度，$\mu_{i\in A}(X)\in[0,1]$，最终选择 $\max(\mu_{i\in A}(X))$ 下的 i 值为 X 的模糊语义值。例如 X 代表"Age"，i 是集合 A 中的一个语义值。$A=\{i \mid i=Young, Middle, Old\}$，$\alpha_i$，$\beta_i$，$\gamma_i$，$\delta_i$ 是梯形模糊隶属函数的参数。由模糊函数可得用于模糊化的规则（如 55≤*realAge*→fuzzyAge ="Old"），并使用 GenericRuleReasoner 进行自动推理。当 *realAge* 为 57 时，自动进行数值模糊化，最终选择"Old"作为 57 岁的语义词。见图 3。

唐旭丽,张斌,傅维刚.情境本体驱动的多源知识融合框架[J].图书情报工作,2018,62(22):109-117.

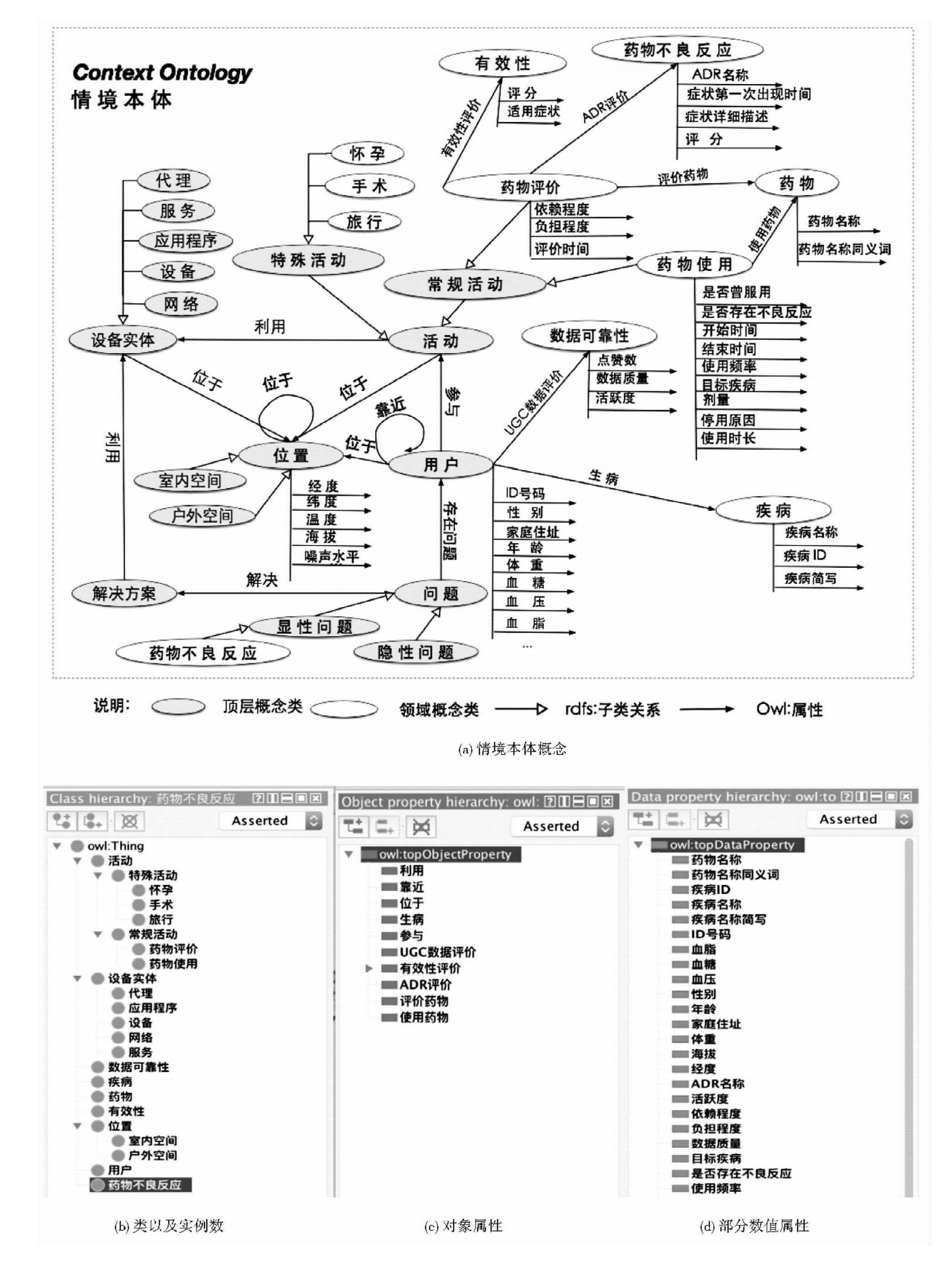


图2 情境本体知识库的构建

3.2.2 本体标准化 将不同格式的本体,如OBO、XML/RDF、OWL/RDF、Turtle、XML、数据库等进行结构统一化是知识融合的基础。一般思路是将数据源统一转换成为OWL/RDF格式[22]。

笔者利用2017年6月版的ADReCS提供的两个本地文件:①ADReCS_ADR_info.xml包含药物ID、药物名称、适应症、药物描述、药物名称的同义词、ADR名称、ADR编码和ADR频率等;②ADReCS_ADR_info.xml包括ADR编码、ADR名称、ADR同义词、产生ADR的药物及药物编码,其中ADR编码通过一组特殊的编码(xx.xx.xx.xxx)来表现药物概念间的层次关系。对于ADReCS,采用的是树形结构的XML文件,而RDF是用一种类似图的结构来表示属性连接资源的陈述,XML文件到RDF的转换不存在万能转换器,笔者根据ADReCS的两个xml文件间的逻辑关联以及ADR编码所表现出的药物概念间的层次关系,抽取出实体Drug、ADR和ADRFrequency,实体间的关系hasADRs、hasADRFrequency、hasDrugs、subclass以及实体各自的对象

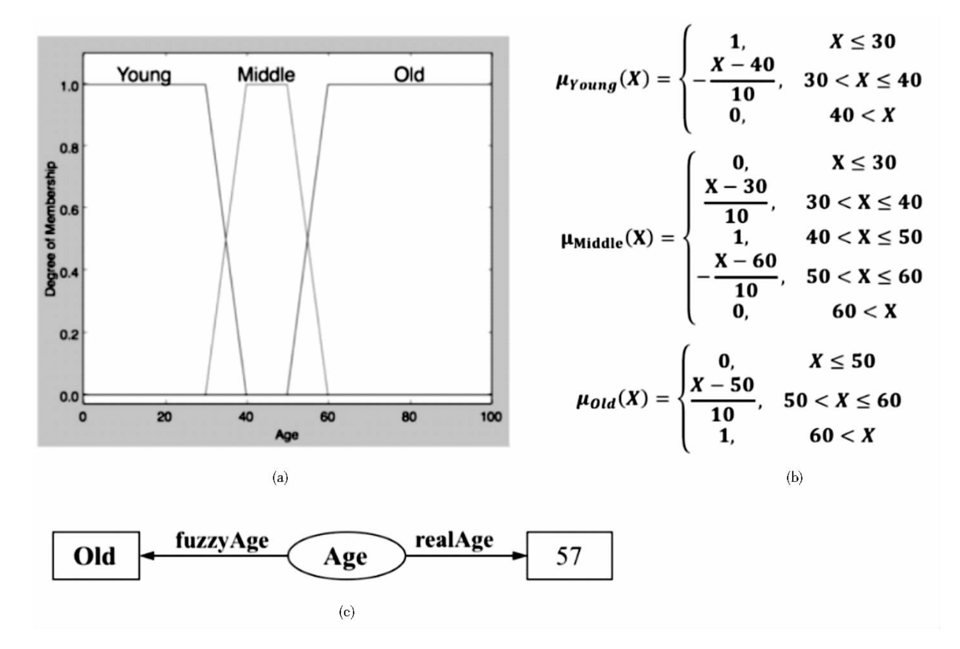


图 3 “Age”的梯形模糊隶属函数

属性,将数据从 XML 文档抽取到 Java 对象中,随后基于该数据使用 Velocity 模版引擎来生成 RDF/XML 文档,并使用 Jena 模型载入 RDF/XML,最终得到 ADReCS. owl。

利用 2017 年 2 月版 Disease Ontology 提供的 doid.owl,包含疾病的名称、类型、病因已经各种疾病概念间的层级关系,Disease Ontology 为 OBO 文件,该 OBO 文件虽然为 RDF 三元组形式,但其内容把所有的节点都视为概念,把疾病实例作为概念层次结构的底层叶子节点。因此需要使用 jena 对该 OBO 文件进行概念的层次遍历,将叶子节点显式转换为该本体的数据层的实例节点。转换后的 Disease Ontology 包含 1 905 个概念节点和 8 276 个实例节点。

3.3 知识融合

在对药物不良反应相关本体进行融合时,本文的桥接点特指本体模式层的类及属性的相等关系,以及本体数据层数据间的相等关系,即不涉及新关系发现的简单融合模式,并在模式层桥接点确定的基础上,使用 Jena 中自带的规则引擎——GenericRuleReasoner,批量实现数据层的链接。

定义 3(模式层融合)设模式层融合的输入为 $F_S = < S_1, S_2, S_3, \cdots, S_m; R_s >$。其中,$S_i$ 表示独立本体 i 的模式层, m 为本体的数目; R_s 表示本体模式层之间的映射关系的集合。$S_i = < C_i, P_i >$,其中 C_i 表示本体 i 中的概念集合; P_i 表示本体 i 中的数值属性和对象属性集合,$i \in \{1, 2, \cdots, m\}$。$R_s$ 中的元素有两种类型:

(1)概念相等的映射关系:(c_i, owl:equivalentClass, c_j) $\in Rs$, $c_i \in C_i$, $c_j \in C_j$, i、j 代表不同的独立本体。

(2)属性相等的映射关系:(p_i, owl:equivalentProperty, p_j) $\in Rs$,其中,$p_i \in P_i$, $p_j \in P_j$, i、j 代表不同的独立本体。

例如本文中 F_S = < ADReCS, Disease Ontology, Context Ontology; R_s >,其中,ADReCS 本体包含概念{药物不良反应(ADR)、药物(drug)、特定类型的药物不良反应频率(ADRFrequency)},情境本体(Context Ontology)包含概念{位置,用户,药物,... 药物评价}。ADReCS 中的药物和情境本体中的药物代表同一个概念,融合时自动为这两个概念添加 equivalentClass 的关系,ADReCS 中的药物概念包含属性{药物名称(DRUG_NAME),药物名称的同义词(DRUG_ SYNONYMS),…}, Context Ontology 中的药物概念包含对象属性{药物名称,药物名称同义词},于是对这两个概念中的相同属性添加映射关系 equivalentProperty,见图 4。

定义 4(数据层融合)设数据层融合的输入为 $F_d = < D_1, D_2, \cdots, D_m; R_d >$。其中,$D_i$ 表示独立本体 i 的数据层, m 为本体数目, R_d 表示本体数据层之间的映射关系的集合。$D_i = < N_i, E_i >$,其中, N_i 表示本体 i 的实例集合,E_i 表示本体 i 的实例关系集合。$E_i \subseteq N_i \times P_i \times N_i$,$P_i$ 表示本体 i 中的数值属性和对象属性集合。R_d 中的元素为实体相等的映射关系:(n_i, owl:sameAs, n_j) $\in R_d$。

例如本文中的情境本体(Context Ontology)中疾病

(disease)概念对应的实例实体 A 的名称 muscular atrophy 和 Disease Ontology 中疾病概念对应的实例实体 B 的名称 muscular atrophy 相等,则为实体 A 和实体 B 添加关系 sameAs,将该规则添加进 Jena 中自带的规则引擎—GenericRuleReasoner,实现自动推理,如图 4 所示:

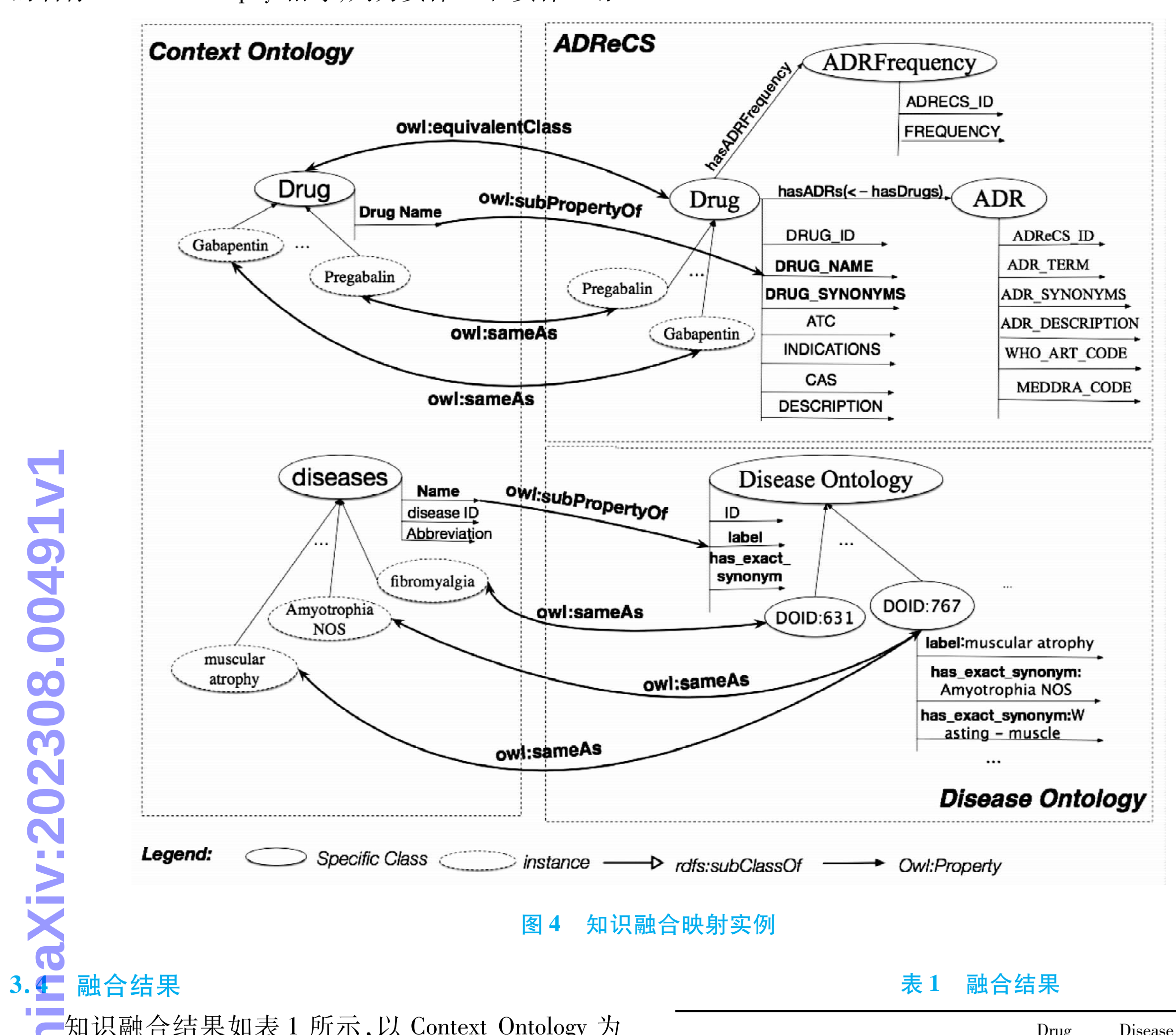


图 4 知识融合映射实例

3.4 融合结果

知识融合结果如表 1 所示,以 Context Ontology 为融合中介,最终形成基于情境的药物不良反应知识库 ConADR Ontology。关于类的融合,实现了 ADRSec Ontology 和 Context Ontology 之间 2 个类(不良反应、药物)的相等关系的映射、Disease Ontology 和 Context Ontology 之间 1 个类(Disease)的相等关系的映射。关于实例的融合,实现了 Disease Ontology 和 Context Ontology 之间 9 个 Drug 实例和 37 个 ADR 实例的相等关系的映射、Disease Ontology 和 Context Ontology 之间 35 个 Disease 实例的相等关系的映射。融合后的 ConADR Ontology 中的类的数目(N)等于融合前的所有本体中类的数目总和(S)减去类映射的数目(F),即 N = S-F;融合后的 ConADR Ontology 中的实例的数目与此同理。

表 1 融合结果

时间	本体知识库	类	Drug实例	Disease实例	ADR实例
融合前	Context Ontology	27	11	38	340
	ADRSec Ontology	3	1 354	0	8 422
	Disease Ontology	1 905	0	8 276	0
融合后	ConADR Ontology	1 932 (2,1)	1 356 (9)	8 279 (35)	8 725 (37)

3.5 案例查询应用

笔者基于 SPARQL 实现了相似案例的推荐,比如一个年纪 37 岁,性别为女,有疾病史纤维肌痛,准备第一次服用 Gabapentin 的患者,想知道自己可能会产生哪些药物不良反应,知识库通过对相似病例进行模糊匹配,给出相似案例的参考结果,见图 5。

4 结语

针对经验相关且非结构化情境下,出现的信息泛滥、信息过载以及难以实现信息按需服务的现状,笔者综合考虑环境情境、个人情境和领域本体三个方面,提出了一种情境本体驱动的知识融合框架。

本文以药物不良反应的知识库构建为例,半自动

Default Graph IRI　http://localhost:8890/DAV/mergedContextOntology

Query

```
PREFIX dis:<http://purl.obolibrary.org/obo/doid.owl#>
SELECT DISTINCT ?DrugName ?Dosage ?ADR_TERM ?ADR_DESCRIPTION
FROM <http://localhost:8890/DAV/mergedContextOntology>
WHERE |
 {
   ?User con:Age 37 .
    con:fibromyalgia con:HasDiseasesBy  ?User.
    ?User con:Gender ?Gender.
    FILTER regex( ?Gender, "F" ).
    ?Drug adr:DRUG_NAME ?DrugName .
    FILTER regex( ?DrugName, "Gabapentin" ) .
    ?DrugUse con:BeEngagedIn ?User.
    ?DrugUse   con:Dosage  ?Dosage .
    ?Drug adr:hasADRs  ?ADR.
    ?ADR adr:ADR_TERM  ?ADR_TERM.
    ?ADR adr:ADR_DESCRIPTION ?ADR_DESCRIPTION.
}
```

Execute　Save　Load　Clear

DrugName	Dosage	ADR_TERM	ADR_DESCRIPTION
Gabapentin	2400 mg	Rash Pustular	Elevated skin lesions associated with the presence of purulent material. [NCI Thesaurus]
Gabapentin	2400 mg	Pain In Extremity	Painful sensation in the upper or lower extremities. [NCI Thesaurus]
Gabapentin	2400 mg	Lacrimation Increased	Not Available
Gabapentin	2400 mg	Headache	A primary headache disorder that is characterized by severe, strictly unilateral PAIN which is orbital, supraorbital, temporal or in any combination of these sites, lasting 15-180 min. occurring 1 to 8 times a day. The attacks are associated with one or more of the following, all of which are ipsilateral: conjunctival injection, lacrimation, nasal congestion, rhinorrhea, facial SWEATING, eyelid EDEMA, and miosis. (International Classification of Headache Disorders, 2nd ed. Cephalalgia 2004: suppl 1) [MeSH]
Gabapentin	2400 mg	Deafness	A general term for the complete loss of the ability to hear from both ears. [MeSH]
Gabapentin	2400 mg	Gastroenteritis	INFLAMMATION of any segment of the GASTROINTESTINAL TRACT from ESOPHAGUS to RECTUM. Causes of gastroenteritis are many including genetic, infection, HYPERSENSITIVITY, drug effects, and CANCER. [MeSH]
Gabapentin	2400 mg	Hyperaesthesia	Not Available

图 5　检索结果

实现了情境本体模式层的构建和数据层的扩充；并以情境本体作为中介本体，实现了情境本体、药物不良反应领域本体 ADReCS 和人类疾病领域本体 Disease Ontology 间的融合；最终在此基础上实现基于 SPARQL 的案例查询。本文的研究的理论框架在其它环境中，如个性化图书馆服务、音乐推荐、高血压服务等方面同样适用，具有可推广性。

本文研究也存在着些不足，因为医药数据的难获取性，数据量比较小，故未在大数据背景下验证其有效性。在后续研究中，笔者会对本文提出的模型框架进行更深入的研究与细化，对本体知识库的关键技术进行优化、改进，以期望发掘其在个性化服务领域的更多应用。

参考文献：

[1] 蒋勋，徐绪堪，苏新宁，等. 知识服务驱动的知识库框架系统内的逻辑架构[J]. 情报理论与实践，2014，37(10)：125－129.

[2] 李枫林，李娜. 基于情景的医药信息服务本体建模及规则推理研究[J]. 情报理论与实践，2016，39(5)：120－124.

[3] 周莉，潘旭伟，谢玉开. 情境感知的电子商务个性化商品信息服务[J]. 图书情报工作，2011，55(10)：130－134.

[4] 潘旭伟，李娜，周莉，等. 情境感知的自适应个性化信息服务体系框架研究[J]. 情报学报，2011，30(5)：514－521.

[5] CANTADOR I，CASTELLS P. Semantic contextualisation in a news recommender system[EB/OL]. [2018－01－10]. https://www. mendeley. com/research-papers/semantic-contextualisation-news-recommender-system/.

[6] REICHLE R，WAGNER M，KHAN M U，et al. A comprehensive context modeling framework for pervasive computing systems[C]// MEIER R，TERZIS S. Ifip wg 6.1 international conference on distributed applications and interoperable systems. Berlin：Springer-verlag，2008：281－295.

[7] SOHN M，JEONG S，LEE H J. Case-based context ontology construction using fuzzy set theory for personalized service in a smart home environment[M]. Berlin：Springer-verlag，2014：1715－1728.

[8] WANG X H，ZHANG D Q，GU T，et al. Ontology based context modeling and reasoning using OWL[C]//IEEE annual conference on pervasive computing and communications workshops. Piscataway：IEEE，2004：18－22.

[9] ALOUI N，GARGOURI F. A generic context ontology for a learning through context－aware annotations memory[C]//JARRAH O A，QUWAIDER M. Omar Al Jarrah and Muhannad Quwaider. Proceedings of the 3rd international conference on information and communication systems. New York：ACM，2012：1－7.

[10] 王丽伟，王伟，高玉堂，等. 领域本体映射的语义互联方法研究——以药物本体为例[J]. 图书情报工作，2013，57(17)：21－25.

[11] 刘晓娟，李广建，化柏林. 知识融合：概念辨析与界说[J]. 图书情报工作，2016，60(13)：13－19.

[12] SMIRNOV A，LEVASHOVA T，SHILOV N. Patterns for context-based knowledge fusion in decision support systems[J]. Information fusion，2015，21(1)：114－129.

[13] 周宇，欧石燕. 面向关联数据的高校机构知识库构建方法研究[J]. 图书情报工作，2016，60(1)：105－113.

[14] PREECE A，HUI K，GRAY A，et al. The KRAFT architecture for knowledge fusion and transformation[J]. Knowledge-based systems，2000，13(2/3)：113－120.

[15] 庄严，李国良，冯建华. 知识库实体对齐技术综述[J]. 计算机研究与发展，2016，53(1)：165－192.

[16] 刘峤，李杨，段宏，等. 知识图谱构建技术综述[J]. 计算机研究与发展，2016，53(3)：582－600.

[17] 王海栋,郑骁庆,张红俊. 基于置信度理论的网络知识融合系统和应用[J]. 计算机系统应用, 2011, 20(1): 1－6.

[18] NIKFARJAM A, SARKER A, O'CONNOR K, et al. Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features[J]. Journal of the American Medical Informatics Association, 2015,22(3):671－681.

[19] BEAN D M, WU H, IQBAL E, et al. Knowledge graph prediction of unknown adverse drug reactions and validation in electronic health records[J]. Scientific reports, 2017,7(1):16416.

[20] 姜兵. 药物不良反应及其监测工作现状[J]. 中国医药科学, 2012,2(24):175－176.

[21] XIE G, XIONG R, CHURCH I. Comparison of kinetics, neural network and fuzzy logic in modelling texture changes of dry peas in long time cooking[J]. LWT- food science and technology, 1998, 31(7/8):639－647.

[22] HEBELER J, FISHER M, BLACE R, et al. Web 3.0 与 Semantic Web 编程[M]. 唐富年,唐荣年,译. 北京:清华大学出版社, 2010.

作者贡献说明:

唐旭丽:论文框架提出与论文撰写;

张斌:论文思路指导及论文修改;

傅维刚:协助论文编程任务和论文修改。

Context Ontology Driven Multi-source Knowledge Fusion Framework

Tang Xuli[1,2]　Zhang Bin[2,3,4]　Fu Weigang[1,2]

[1] Center for Studies of Information Resources, Wuhan University, Wuhan 430072

[2] Big Data Institute, Wuhan University, Wuhan 430072

[3] National Institute of Cultural Development, Wuhan University, Wuhan 430072

[4] Center of Traditional Chinese Cultural Studies, Wuhan University, Wuhan 430072

Abstract: [**Purpose/significance**] Context-awareness modeling is an important method to solve information over-flow, information overload, and to realize information on demand, however, it always being ignored in the construction of knowledge base, which hinders the practical application of knowledge base as well as reduces the efficiency and effectiveness of knowledge service. [**Method/process**] This paper proposed an ontology-based context driven multi-source knowledge fusion framework taking the context, personal profiles and domain ontology into consideration. Under the guidance of this framework, this paper constructed an Adverse Drug Reactions (ADR) knowledge base with respect to the contextual relevance naming ConADR Ontology. Firstly, we constructed a situation ontology which can semi-automatically update schema and extend ontology instance, and then successfully fuse it with existed domain ontology ADReCS and Disease Ontology using Jena and Protégé. Finally, we developed a case query application based on SPARQL. [**Result/conclusion**] The example shows that the framework has a certain feasibility and theoretical reference value for the merger and construction of knowledge base.

Keywords: context ontology　multi-source knowledge fusion　construction of knowledge base　adverse drug reactions